Change, Change, Change: three approaches

Tom Costello
Computer Science Department
Stanford University
Stanford, CA 94305
email: costelloQcs.Stanford.EDU

Abstract

We consider the *frame problem,* that is, characterizing the assumption that properties tend to persist over time. We show that there are at least three distinct assumptions that can be made. We show the first assumption, which have been widely studied, is not naturally captured by *circumscription.* The first assumption is, "there is as little change as possible between one situation and the next". This is closely related to temporal projection.

The second assumption is that actions have as few effects as possible. This has arisen in causal approaches. We show this assumption cannot be captured by any circumscription policy, as it compares models with different domains.

We consider a third assumption—there are as many frame axioms true as possible—which can be captured by circumscription. All three assumptions differ, which we show by giving examples.

All agree on a small class of theories, those axiomatized by effect axioms and a class of sentences we call *compatible binary domain constraints.* Further, for a similar class of theories a deduction theorem holds, allowing observations, or facts about particular situations, to be brought through the non-monotonic consequence relation. This justifies approaches based on separating "observations" from "effects", and applying projection to the effects, to solve problems based on causality reasoning.

1 The Frame Problem

Reasoning about change is central to much of common sense reasoning. To reason about change we clearly need to represent what changes after an action is completed. What is not so immediately obvious is that we also need to allow the reasoner to infer what does not change after an action. The simplest way to specify what changes and what does not would be to write each set down. However this seems wasteful, and worse, seems likely to introduce inconsistencies. It seems clear that just specifying what changes is sufficient—everything else doesn't change. We can generate the properties that do not change from the list of properties that do change. If we do this we are guaranteed that our lists will be consistent. Moreover, usually fewer things change than don't change, and thus we will have much less to specify.

The idea that we need specify just the positive instances of change has given rise to a problem: How do we generate all the instances of properties that stay the same? This is how I formulate the *frame problem.*

Since its inception the frame problem has mutated. Some people view it as the problem of correctly determining the consequences of a set of sentences describing some changing domain. This might include concurrent, nondeterministic, under-specified or continuous actions. We do not deal with this extended notion here.

We do keep in mind one modification to the original statement of the frame problem. We do not just want to encode the things that change and don't change compactly, we also want to be able to add new changes by adding new sentences. This requirement forces us to use some non-monotonic reasoning. This requirement is called *elaboration t*olerance[Costello, 1997a] by McCarthy[l992].

1.1 The Situation Calculus

To look at the frame problem we need a logical language to represent time, events, and the properties that hold in the world. We propose to use the situation calculus, a formalism for reasoning about events in the world. The paper that introduced situations [McCarthy and Hayes, 1969] defined them as follows, "A situation 5 is the complete state of the universe at an instant of time". The basic mechanism used in the calculus to define a new situation is the result function,

$$s' = \mathbf{result}(e, s).$$

In this formula s is a situation, and e is an event, and s^f is a new situation that results when e occurs. A function or predicate of a situation is called a *fluent.*

1.2 Frame Axioms

As we mentioned earlier, a certain problem arises when we formalize domains in this language. Assume there is an axiom that states what happens to the location of a block when an action *move* is performed.

$$\forall s.\mathrm{holds}(\mathrm{clear}(l), s) \rightarrow \mathrm{holds}(\mathrm{at}(x, l), \mathrm{result}(\mathrm{move}(x, l), s))$$

Sentences of this form are called *effect axioms.* If another property, *colour,* is added, the axiom which described what effect the action had on the location does not give us any information on what happens to the new fluent *colour.* The usual intent is that the new fluent remains unchanged by the action, but this is not always the case. To specify that the colour of the block does not change we must write:

$$\forall x\ c\ s.\mathrm{holds}(\mathrm{colour}(x, c), s) \rightarrow \mathrm{holds}(\mathrm{colour}(x, c), \mathrm{result}(\mathrm{move}(x, l), s))$$

To express all the fluents that do not change, an axiom would be needed for each action/fluent pair. These axioms stating that fluents do not change are called *frame axioms.* Having to specify all this lack of change seems extravagant, and more importantly it seems unnecessary to have to write these out individually. It seems that it should be possible to generate these frame axioms automatically. The *frame problem* is the problem of characterizing these frame axioms.

1.3 Plan of Paper

We first define the situation calculus and circumscription, the language and non-monotonic framework we consider. A fuller exposition of these can be found in [McCarthy and Hayes, 1969] and [McCarthy, 1986]. We then consider the *projection problem.* We give the preference on models that it suggests, and show that it is not naturally captured by circumscription. We show it is captured by a II{ formula, (that is a formula all of whose second order quantifiers are universal) which is not in the form of circumscription.

Next we consider the *causality assumption,* that actions have as few effects as possible. We show that this cannot be captured by any circumscription policy.

We consider the "frame" assumption, that there are as many frame axioms as possible. We show that this can be captured by a form of circumscription that allows free predicate variables.

We then consider the various other proposals for capturing these defaults that have been suggested. We show that causal reasoning, as proposed by Gelfond and Lifschitz[l993] corresponds to the *causality assumption.* We show that for a restricted class of theories, *chronological minimization* corresponds to projection, and present a new counter-example, a simple theory, with only facts about the initial situation and disjunction of effects axioms, where *chronological minimization* does not give the correct results.

We show that for the restricted class where chronological minimization applies, an explanation closure axiom is validated. This allows a deduction theorem which we can state as follows. Let $\models_{proj}$ give the consequences of the projection assumption, and $\models_{caus}$, the consequences of the causality assumption, that is minimal entailment under $\leq_{proj}$ and $\leq_{caus}$, and let Γ be a theory axiomatized by effects axioms, $t, t_1, \ldots, t_i$ be constant situation terms, $F, f_1, \ldots, f_n$ be fluent constants, for $n \in \omega$,

$$\Gamma \models_{proj} \bigwedge_{i \leq n} \mathrm{holds}(f_i, t_i) \rightarrow \mathrm{holds}(F, t)$$
if and only if
$$\Gamma, \{\mathrm{holds}(f_i, t_i) | i \leq n\} \models_{caus} \mathrm{holds}(F, t)$$

This justifies the approaches of Sandewall, Lifschitz and others that have solved causal reasoning problems by dividing a theory in two, and applying projection type reasoning to a part, and re-conjoining the reserved sentences.

2 The Situation Calculus

The situation calculus of order n, L_{sit}, is a language of n^{th} order logic with equality, with three disjoint sorts: fluent names, actions, and situations[1]. It contains exactly one predicate constant, holds, in addition to equality, one function constant result(a, *s)* which returns a situation, a function sit(g) from predicates of fluents to situations, and one situation constant SO, plus a set of fluent constants $\overline{f}$, and a set of action constants $\overline{a}$. When we wish to specify the action and fluent constants we write $L_{sit}(\overline{f}, \overline{a})$. We shall sometimes add a predicate ab(a,f,s).

3 Circumscription

Non-monotonic logics are logics where the consequences of a set of sentences are not necessarily consequences of supersets of that set of sentences.

This paper uses circumscription as its non-monotonic machinery. Circumscription is a form of non-monotonic logic introduced by McCarthy [1986]. It expresses the non-monotonic consequences of a finitely axiomatizable theory *A* in a language L, under a certain circumscriptive policy, as a sentence of second order logic. A circumscriptive policy is a choice of a finite set of formulas of *L* to minimize

4 The Projection Problem

The *projection assumption* is the default that there are as few changes from one situation to the next as possible. The critical change between this assumption, and the naive approaches that fall victim to the Yale Shooting Problem, is that situations that differ on what fluents hold, before the action we consider, are not compared.

When we consider the set of changes that happen when an action occurs, it would not make sense if the two situations from which we measure have different properties.

[1] We use *s* with primes and indices to range over situations, and *a* and / to range over actions and fluent names, and *g* to range over unary predicates of fluents.

Definition: 1
The ordering on models that models the assumption of the projection problem is defined as follows. Let $\mathfrak{A}, \mathfrak{B}$ be two models of $L_{sit}(\bar{a}, \bar{f})$. We say $\mathfrak{A} \leq_{proj} \mathfrak{B}$, if for all situations s and $R(a,s) = \mathfrak{A}[\text{result}(a,s)]$ and $R'(a,s) = \mathfrak{B}[\text{result}(a,s)]$, where $a \in \bar{a}$, then

$$\begin{array}{l} \text{If } \forall f \in \bar{f}.\langle f,s\rangle \in \mathfrak{A}[\text{holds}] \text{ iff } \langle f,s\rangle \in \mathfrak{B}[\text{holds}] \text{ then} \\ \forall a \in \bar{a}. \\ \{f | \langle f,s\rangle \in \mathfrak{A}[\text{holds}] \not\equiv \langle f, R(a,s)\rangle \in \mathfrak{A}[\text{holds}]\} \subset \\ \quad \{f | \langle f,s\rangle \in \mathfrak{B}[\text{holds}] \not\equiv \langle f, R'(a,s)\rangle \in \mathfrak{B}[\text{holds}]\} \end{array}$$

A possible criticism of this approach is that it does not count the extra changes that might be implied by domain constraints. This criticism is invalid, as for each other change that is implied by a domain constraint, there is another effect axiom that can be derived. Because we have stayed in the language of the situation calculus, we have avoided the problems with domain constraints that plague approaches that introduce new predicates.

Theorem: 1
The assumption of the projection problem is expressible in the 11} fragment of second order logic.

Proof: The following sentence is the required Π^1_1 sentence.

$$\begin{array}{l} A[\text{holds}, ab] \wedge \forall \text{holds}' ab'. \\ \quad (A[\text{holds}', ab'] \wedge \\ \quad (\forall a, f, s.(\forall f'.\text{holds}(f',s) \equiv \text{holds}'(f',s))) \rightarrow \\ \quad\quad (ab(a,f,s) \rightarrow ab'(a,f,s)))) \rightarrow \\ \quad \forall a, f, s.(ab'(a,f,s) \equiv ab(a,f,s)) \blacksquare \end{array}$$

There is no circumscription policy that captures the projection assumption in this language. This is shown in [Costello, 1997b], by showing that the filters of the partial order $<_{proj}$ are not finitely first order axiomatizable.

However, if we extend our definition of circumscription to allow arbitrary sentences to describe the preference on models, then we can easily describe this preference. Lifschitz [1984] suggests the name *general circumscription* for minimal entailment where the preference on models is described by a first order formula.

Giunchiglia

Giunchiglia [1996] suggests a second order sentence, that is logically equivalent (modulo translation into a variant of the situation calculus) to the projection assumption we consider. He considers the same ordering on models, only comparing situations that agree on all fluents. This ordering on models has been considered by Lin and Reiter[l994]. They do not characterize it in second order logic, rather they use a two stage translation with predicate completion.

5 The Causality Problem

Rather than minimize changes from one situation to the next, another approach is to minimize the effects that actions have. That is, we assume that all change is caused, and then minimize the effects that each action causes.

To state that an action causes a change, we write an effect axiom, that is a sentence of the form,

$$\begin{array}{l} \forall s. \bigwedge_{f' \in P} \text{holds}(f',s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s) \rightarrow \\ \quad (\text{holds}(f,s) \equiv \neg\text{holds}(f, \text{result}(a,s))) \end{array}$$

for some particular P and N.

This states that if the fluents P hold, and the fluents N do not, then the action a causes f to not hold. This can be written in other ways, [Gelfond and Lifschitz, 1993] for instance, a **causes** $\neg f$ **if** $\{f' | f \in P\}, \{\neg f'' | f \in, N\}$, or $cancels(a, f, P, N)$. The meaning of these is intuitively the same as the effect axiom.

To state that all changes are caused we write an explanation closure axiom.

$$\begin{array}{l} \forall a, f, s.\text{holds}(f, \text{result}(a,s)) \equiv \\ \quad \exists N, P.(\forall s' \bigwedge_{f' \in P} \text{holds}(f',s') \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s') \rightarrow \\ \quad\quad (\neg\text{holds}(f,s') \rightarrow \text{holds}(f, \text{result}(a,s')))) \wedge \\ \quad \bigwedge_{f' \in P} \text{holds}(f',s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s) \wedge \neg\text{holds}(f,s) \\ \vee \\ \quad \text{holds}(f,s) \wedge \\ \quad \forall N, P.\neg(\forall s'. \bigwedge_{f' \in P} \text{holds}(f',s') \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s') \rightarrow \\ \quad\quad (\text{holds}(f,s') \rightarrow \neg\text{holds}(f, \text{result}(a,s')))) \vee \\ \quad \neg(\bigwedge_{f' \in P} \text{holds}(f',s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s)) \end{array}$$

We note that this is equivalent to,

$$\begin{array}{l} \forall a, f, s, s'.(\forall f'.\text{holds}(f',s) \equiv \text{holds}(f',s')) \rightarrow \\ \quad (\text{holds}(f, \text{result}(a,s)) \rightarrow \text{holds}(f, \text{result}(a,s'))). \end{array}$$

That is, in this case it states that if a change occurs in a situation then it occurs in every situation where the same fluents hold. We shall later show that this axiom is naturally satisfied if there are no "observations".

Definition: 2
The ordering on models that models the assumption of the causality problem is defined as follows. We prefer models where actions have fewer effects. We can measure the effects of an action by the set of formulas of the forms:

$$\begin{array}{l} \forall s. \bigwedge_{f' \in P} \text{holds}(f',s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'',s) \rightarrow \\ \quad (\text{holds}(f,s) \equiv \neg\text{holds}(f, \text{result}(a,s))) \end{array}$$

Let ϕ range over formulas of these forms. Let $\mathfrak{A}, \mathfrak{B}$ be two models of L_{sit}. We say $\mathfrak{A} \leq_{caus} \mathfrak{B}$, if

$$\mathfrak{B} \models \phi \text{ implies } \mathfrak{A} \models \phi$$

We note that these models are assumed to obey the rule of explanation closure above.

Theorem: 2
The assumption of the causality problem is not expressible in circumscription.

Proof:
Circumscription does not compare models of differing cardinality, but to minimize universals it is sometimes necessary to enlarge the domain, adding a new object that witnesses the existential.

Consider the theory in $L_{sit}(\{A\},\{F\})$, axiomatized by holds(F, 50) The non-monotonic consequences of this should include

$$\text{holds}(F, S0) \wedge (\forall s.\text{holds}(F, s) \rightarrow \text{holds}(F, \text{result}(A, s)))$$

However, in any model where all situations are denoted by closed terms, the following will be satisfied,

$$\forall s.\neg\text{holds}(F, s) \rightarrow \text{holds}(F, \text{result}(A, s)).$$

Thus, to deny this effect axiom, we need to add a situation not denoted by any closed term. I

6 The "Frame" Problem

We now consider another assumption, one which to our knowledge has not been previously considered. We prefer models where there are more frame axioms true. This seems like a very natural approach. We first consider how to represent this assumption.

Definition: 3
The ordering on models that models the assumption of the frame problem is defined as follows. We prefer models where there are more frame axioms true. Frame formulas are of the form, where N and P are sets of fluent constants:

$$\forall s. \bigwedge_{f' \in P} \text{holds}(f', s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'', s) \rightarrow (\text{holds}(f, s) \equiv \text{holds}(f, \text{result}(a, s)))$$

Let ϕ range over formulas of these forms. Let $\mathfrak{A}, \mathfrak{B}$ be two models of L_{sit}. We say $\mathfrak{B} \leq_{frame} \mathfrak{A}$, if

$$\mathfrak{B} \models \phi \text{ implies } \mathfrak{A} \models \phi$$

Theorem: 3
The assumption of the frame problem is not expressible in circumscription.

Proof: No finite set of first order sentences describes the same order on models as the infinite set of formulas described above.

Theorem: 4
The assumption of the frame problem is expressible in circumscription, if we allow the minimized formulas to contain second order variables.

Proof: If we allow a unary predicate variable for fluents g the class of formulas above can be parameterized by g, as:

$$\forall s.(\forall f'.g(f') \equiv \text{holds}(f', s)) \rightarrow (\text{holds}(f, s) \equiv \text{holds}(f, \text{result}(a, s)))$$

The circumscription formula is then,

$$\begin{aligned}
&A[\text{holds}] \wedge \forall \text{holds}'.[A[\text{holds}'] \wedge \\
&\quad \forall a, f, g.(\forall s.(\forall f'.g(f') \equiv \text{holds}'(f', s)) \rightarrow \\
&\qquad (\text{holds}'(f, s) \equiv \text{holds}'(f, \text{result}(a, s)))) \rightarrow \\
&\qquad (\forall s.(\forall f'.g(f') \equiv \text{holds}(f', s)) \rightarrow \\
&\qquad\quad (\text{holds}(f, s) \equiv \text{holds}(f, \text{result}(a, s))))] \rightarrow \\
&(\forall f, s.\text{holds}(f, s) \equiv \text{holds}'(f, s))
\end{aligned}$$

■

To the best of the author's knowledge this approach, of preferring models where there are more frame axioms has not been previously considered.

7 Examples

We show, via examples, that each approach is distinct.

Example: 1
The first example concerns reasoning forwards in time. For simplicity we consider a language with only one action A, and one fluent-name F.

$$\begin{array}{l} \text{holds}(F, S0) \\ \forall s.\text{holds}(F, s) \rightarrow \text{holds}(F, \text{result}(A, S0)) \end{array} \quad (1)$$

We consider that a conclusion that should be drawn from this theory, if we assume the causality assumption is,

$$1 \models_{caus} \neg\forall s.\neg\text{holds}(F, s) \rightarrow \text{holds}(F, \text{result}(A, s))$$

In contrast the "frame" assumption will not entail this, as this does not add a new frame axiom. The only (nontrivial) possible frame axioms are,

$$\begin{array}{l} \forall s.\text{holds}(F, s) \rightarrow \text{holds}(f, s) \\ \forall s.\neg\text{holds}(F, s) \rightarrow \neg\text{holds}(f, s) \end{array}$$

Thus the frame assumption will add exactly these two sentences, which do not entail the negation of the effect axioms above.

We mention for completeness that the standard example for the distinction between projection and causality is the Stanford murder mystery, where both the causality and frame assumption give that the gun was loaded initially, while the projection assumption does not.

Example: 2
The first example concerns reasoning backwards in time. For simplicity we consider a language with only one action A, and two fluent-names F_1, F_2. We are told that F_1 holds at the second situation, result(A, 50). We are also told that $F\backslash$ does not hold at the first situation. Finally we are told that A causes $F\backslash$ is F2 holds.

$$\neg\text{holds}(F_1, S0), \quad \text{holds}(F_1, \text{result}(A, S0))$$
$$\forall s.\text{holds}(F_2, s) \rightarrow \text{holds}(F_1, \text{result}(A, s))$$

The result of the projection assumption adds,

$$\forall s.\text{holds}(F_1, \text{result}(A, S0)) \rightarrow$$
$$\text{holds}(F_1, \text{result}(A, \text{result}(A, S0))),$$

and $\forall s.\text{holds}(F_2, s) \equiv \text{holds}(F_2, \text{result}(A, s))$, while the frame and causality assumptions both give, $\text{holds}(F_2, S0)$ in addition.

The causality assumption also gives the negation of the other effect axiom.

The causality problem has been addressed by Lin and Reiter[l994], Kartha and Lifschitz[l995] and Sandewall[l994] among others. We consider their approaches, and show how they attempt to avoid the result of the previous theorems.

8 Nested Abnormality and Filter entailment

Lifschitz and Sandewall suggest extensions to circumscription, or non-monotonic systems that are based around minimal entailment. Sandewall's systems [1994] are based on filter preferential entailment. Kartha and Lifschitz [1995] bases their system on nested abnormality theories[Lifschitz, 1995]. Both these formalisms have the property that they necessarily divide the theory they are non-monotonically completing into subsets, based on some syntactic property. In particular they divide observations from causal rules or effect axioms. However, if a causal rule is disjoined with and effect axiom this cannot be achieved. They then solve the projection problem, and conjoin observations to solve the causality problem.

Preferential Entailment

Sandewall uses the term *filter preferential entailment* to denote a type of semantic entailment. He imagines a case where we have a theory divided into two parts, *T* and To. He divides the theory into two sets of sentences, so he can apply a selection function to one set, and then conjoin the other set to get the final theory. If ζ is a selection function, that is it maps a set of models to a subset of those models, then the non-monotonic consequences of Γ and Γ_0 are $\Gamma \cup \zeta(\Gamma_0)$.

The primary intuition behind Sandewall's preference on models is similar in spirit to chronological minimization. Changes occur preferentially later in time. We now show that our projection assumption can sometimes give the same results as chronological minimization.

In the case of temporal projection, when we do not have facts about the initial situation, and where we have only effect axioms and domain constraints, the same results can be achieved by not comparing situations that differ, as we earlier propose. Not comparing situations that differ on what is the case seems obvious. Apples should not be compared with oranges.

A domain constraint is a sentence in L_{sit} whose only situation terms are universally quantified variables. A binary domain constraint is a formula of the form, $\forall s.(\neg)\text{holds}(f, s) \rightarrow (\neg)\text{holds}(f', s)$, for fluent constants / and /'.

Theorem: 5
Given a theory T in L_{sit}, axiomatized by binary domain constraints and effect axioms, which do not entail any domain constraint not equivalent to a conjunction of binary domain constraints, then the result of the projection assumption is equivalent to putting off change as long as possible (chronological minimization).

Proof: We first note that the axiom of explanation closure is validated under both assumptions.

Secondly we see that for a given model $\mathfrak{M}$ and a situation *s,* by explanation closure, we can determine what holds in the following situations by looking that the initial situation, in some model $\mathfrak{M}'$, where the fluents that initially hold in $\mathfrak{M}'$ are the same as hold at situation *s* in m. As we have only effect axioms and binary domain constraints, there will always exist such a model.

Thus it suffices to look at only one step of projection. However, when we just do one step, we minimize change in both cases. ∎

Nested Abnormality Theories

Lifschitz in [1995] suggest a very similar idea to Sandewall's. Rather than describing the idea in semantic terms, Lifschitz uses the notation of circumscription.

He divides his theory into several parts. To each part he applies a circumscription. He does this uniformly by defining the predicate *ab,* of an appropriate arity to be predicate to be minimized. All predicates are varied by default.

"Deduction Theorems"

To a large extent both of these approaches attempt to solve the problem of causality by solving the projection problem, and then adding back certain facts. This is why they need to be able to syntactically divide theories into two sub-theories.

We now prove a theorem showing that solving the projection problem for theories without observations, and adding observations is equivalent, for closed situation terms, to solving the causality problem. A similar theorem can be established for the frame assumption.

Theorem: 6
Let T be a theory in L_{sit}, that is axiomatized by effect axioms and binary domain constraints, where every domain constraint in T is equivalent to a conjunction of binary constraints. Then, if t is a constant situation term, and $t_1, \ldots, t_n$, $F, f_1, \ldots, F_n$, $n \in \omega$ are constant situation terms and fluents,

$$T \models_{proj} \bigwedge_{i \leq n} \text{holds}(f_i, t_i) \rightarrow \text{holds}(f, t)$$
if and only if
$$T, \{\text{holds}(f_i, t_i) | i \in n\} \models_{caus} \text{holds}(f, t)$$

Proof

All consistent theories T axiomatized by effect axioms and binary domain constraints are consistent with the statement of explanation closure, subject to the above proviso. As minimal entailment is cumulative, thus we may add explanation closure to the left hand side.

We then prove the above claim by induction on the size of the *Vs.* If they are all S'O, then the implication its true only if it is an instantiation of a binary domain constraint, and thus the equivalence holds.

But, then we see that $(\forall f.\mathrm{holds}(f,t) \equiv g(f)) \wedge \neg\mathrm{holds}(F,t) \rightarrow \mathrm{holds}(F,\mathrm{result}(a,t))$, is equivalent to $\forall s.(\forall f.\mathrm{holds}(f,s) \equiv g(f)) \wedge \neg\mathrm{holds}(F,s) \rightarrow \mathrm{holds}(F,\mathrm{result}(a,s))$, explanation closure. This gives us the induction step to prove the above claim. I

Thus the approaches of Kartha and Lifschitz, and Sandewall, which claim to do causal reasoning, are in fact doing temporal projection. They simulate causal reasoning because for the class of problems they consider, the above theorem applies.

8.1 Lin and Shoham's Approach

This approach quantifies over circumscription. They take the consequences of a theory to be the intersection of the consequences of a set of circumscriptions. Lin and Shoham [Lin and Shoham, 1995] suggest a solution that again returned to minimizing the total amount of change, but they suggested that this should be minimized one situation at a time. For every situation, and possible set of fluents that could hold at that situation, they calculate what the minimal amount of change between that situation and its successors could be. They then add the conditional that if those fluents held at the situation, then the successors will be exactly those that have minimal differences. The major drawback with this approach is that it can become inconsistent when given disjunctive information. However, Lin and Shoham show that it behaves correctly for a limited class of theories, essentially those where what holds in the initial situation uniquely determines what is holds in all later situations.

9 Conclusion

We considered three distinct defaults about temporal reasoning. The first, *projection* has been considered before by Lin and Reiter, and Giunchiglia. The second, minimizing the effect axioms that hold, corresponds to causal reasoning. This has been considered semantically by many researchers. We show for the first time that this cannot be captured by circumscription, as it compare models with different cardinalities. The third method, maximizing the set of frame axioms is novel, but perhaps is the most natural.

We established deduction type relationships between these approaches, explaining why current approaches that divide theories into parts behave as they do. We gave conditions for when applying projection reasoning, then adding observations, is equivalent to causality reasoning.

A more comprehensive survey is given in [Costello, 1997b]. However, to the best of the author's knowledge, all previous approaches either attempted to minimize change (chronologically), or minimize *causality.*

10 Acknowledgments

This research is supported in part by ARPA/Rome Laboratory planning initiative under grant (ONR) N00014-94-1-0775. I am grateful to John McCarthy and Anna Patterson who gave useful criticism on earlier drafts.

References

[Costello, 1997a] Tom Costello. Beyond Minimizing Change. In *Proc. National Conference on Artificial Intelligence (AAAI '97),* 1997.

[Costello, 1997b] T bm Costello. *Non-monotonicity and Change.* PhD thesis, Stanford University, 1997.

[Gelfond and Lifschitz, 1993] M. Gelfond and V. Lifschitz. Representing Actions and Change by Logic Programs. *Journal of Logic Programming,* 17:301-322, 1993.

[Giunchiglia, 1996] E. Giunchiglia. Determing Ramifications in the Situation Calculus. In J. Doyle, editor, *Proceedings of KR 1996.* KR, 1996.

[Kartha and Lifschitz, 1995] G. N. Kartha and V. Lifschitz. A simple formalization of actions using circumscription. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCA1 '95),* 1995.

[Lifschitz, 1984] V. Lifschitz. Some Results on Circumscription. In *First Workshop on Non-monotonic Reasoning,* pages 151-164, Stanford University, August 1984/

[Lifschitz, 1995] V. Lifschitz. Nested abnormality theories. *Artificial Intelligence,* 74:351-365, 1995.

[Lin and Reiter, 1994] F. Lin and R. Reiter. State Constraints Revisited. *Journal of Logic and Computation,* 4(5):655-678, 1994.

[Lin and Shoham, 1995] F. Lin and Y. Shoham. Provably Correct Theories of Action. *Journal of the ACM,* 42(2):293-320, 1995.

[McCarthy and Hayes, 1969] J. McCarthy and P. Hayes. Some Philosophical Problems From the Standpoint of Artificial Intelligence. In D. Michie, editor, *Machine Intelligence 4,* pages 463-502. Edinburgh University Press, Edinburgh, UK, 1969.

[McCarthy, 1986] J. McCarthy. Applications of Circumscription to Formalizing Commonsense Knowledge. *Artificial Intelligence,* 28:89-116, 1986.

[McCarthy, 1992] John McCarthy. Overcoming and Unexpected Obstacle. Available as http://www-formal.stanford.edu/jmc/elaboration.html, 1992.

[Sandewall, 1994] E. Sandewall. *Features and Fluents.* Oxford University Press, 1994.

TEMPORAL REASONING

Temporal Reasoning 2